**Facilitated Access and Application of Computational
Astrostatistics Algorithms
Annual Report for Grant: NNG08EJ50I**
PI: Chris Miller
August 1, 2008 to July 31, 2009 (Year 1 of 3)

CTIO - National Optical Astronomy Observatory
**Colina El Pino s/n Casilla 603 La Serena Chile
950 N. Cherry Ave. Tucson, AZ 85719 US**

Project Summary:

The goal of this project is to build a set of well-documented libraries and web-services for computational and statistical algorithms designed for use by astronomers. These algorithms are developed by the International Computational Astro-statistics group (InCA), and include active learning search algorithms, KD-tree codes, fast mixture model algorithms, non-parametric statistics and more.

We recognize the need in the astronomical community for a well-documented library of these tools. This project will provide that library in a variety of scripting languages and via web services.

Summary of Accomplishments:

According to the Project Time-line in the proposal, Year 1 was to be spent developing IDL, Python, and R interfaces for existing InCA codes. In order to accomplish this, we had planned on hiring a post-doctoral researcher to lead this effort. After the NASA funds arrived at the NOAO, a job advertisement was immediately placed in the AAS Job Register (#24604) in April, 2008. Unfortunately, we were unable to hire a post-doc due to the lack of qualified applicants.

In order to keep the project on track, The PI (Miller) and the Co-I (Schneider-CMU) decided to work with a number of graduate students at CMU. Thus, Joey Richards (CMU Statistics PhD student) came to NOAO-South to work directly with the PI on the new InCA-developed diffusion mapping algorithms. Richards has provided the diffusion mapping libraries in both Python and IDL. We are currently applying those libraries in a "real-world" test. The diffusion mapping technique is being used to model the star-formation histories of all SDSS DR7 galaxy spectra. The Python libraries have been deployed onto a Condor high performance 24-node cluster to accomplish this task. The effort is fully documented in such a way that other astronomers will be able to do similar research on their own Condor networks.

We also worked with Tzu-Kuo Huang (CMU Computer Science PhD Student) to develop dynamical modeling and prediction algorithms. Tzu-Kuo and Co-I Schneider are developing algorithms to model and predict dynamical systems when no temporal information is available. For example, this technique can be used to predict the movement of stars within a galaxy (spiral disk or elliptical infall), or to predict the evolution of stars along an HR-diagram. This type of time-domain science is becoming more important with the advent of new time-domain wide field surveys like the LSST. Thus, we coordinated closely with Jeff Kantor and Andrew Connolly (LSST-Corp and Univ. Of Washingon) on this effort such that we can build these applications into LSST science efforts.

The PI (Miller) started a collaboration with the Center for Mathematical Modeling (CMM) at the Universidad de Chile in Santiago Chile. Our goal is to work more closely with scientists from CMM to enable the use of InCA algorithms within the environment of Chilean astronomy (which will include the LSST). The PI (Miller) hired a short-term undergraduate researcher from CMM (Mauricio Hidalgo) to begin work with porting the InCA non-parametric fitting codes from IDL into Python. Progress was made and we expect the libraries to be available soon.

Finally, towards the end of the reporting period, we were able to hire Craig Harrison (PhD) into the project as a post-doctoral researcher. Harrison has a solid background in statistics and computer science as well as in astronomy. He is a perfect fit for the project. Craig will begin work with the PI (Miller) at NOAO-South in August, 2009.

To summarize our Year 1 efforts, we are off to a slow but steady start to the project. Despite our failure to hire the required post-doctoral researcher in Year 1, we used our AISR funds to make progress on the InCA library development by leveraging our connections to the strong computer-science and statistics graduate programs at CMU. The PI was also able to leverage a new connection with the Center for Mathematical Modeling in Santiago, Chile. These connections will prove fruitful over the timeline of the project, in that we now have dedicated students working towards goals defined in the AISR proposal. In addition, we have succesfully hired a post-doctoral researcher who will be able to lead the project into Years 2 and 3.

Talks and or Publications:

"Learning Linear Dynamical Systems without Sequence Information" Tzu-Kuo Huang, Jeff Schneider, ICML 2009: Proceedings of the 26th International Conference on Machine Learning

Invited talk: Chris Miller at the Center for Mathematical Modeling, Univ. Chile, Santiago Chile: "**Facilitated Access and Application of Computational Astrostatistics Algorithms", May 12, 2009**

Invited talk: Joey Richards, Center for Mathematical Modeling, Univ. Chile, Santiago Chile: "**Diffusion Mapping Techniques to measure the Star-formation histories of galaxies"**, May 12, 2009